

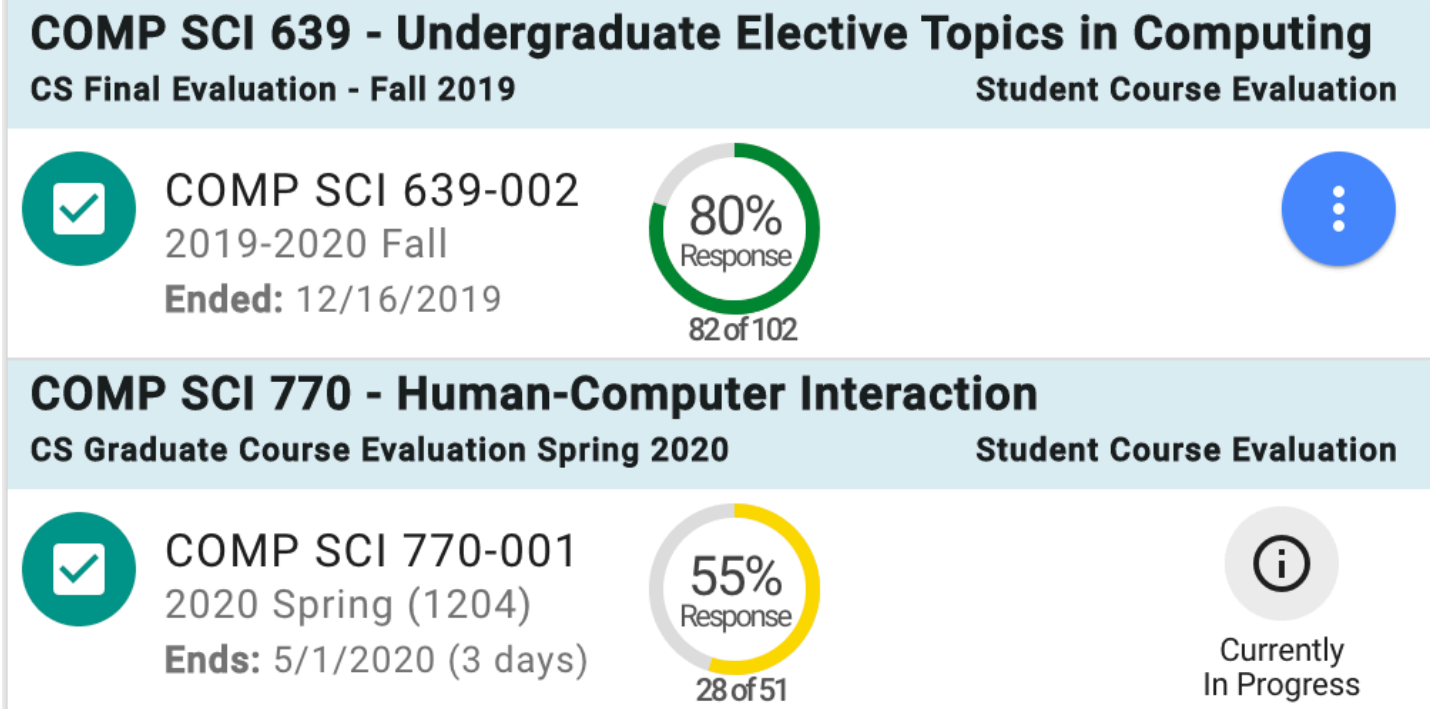
Human-Computer Interaction

# Reporting & Writing HCI Papers

Professor Bilge Mutlu

# Announcements

- >> Today is our *last* class
- >> Please complete the course evaluation (through AEFIS) by **May 1**
- >> Response rate is currently at **55%**; last semester was **80%**



# Today's Agenda

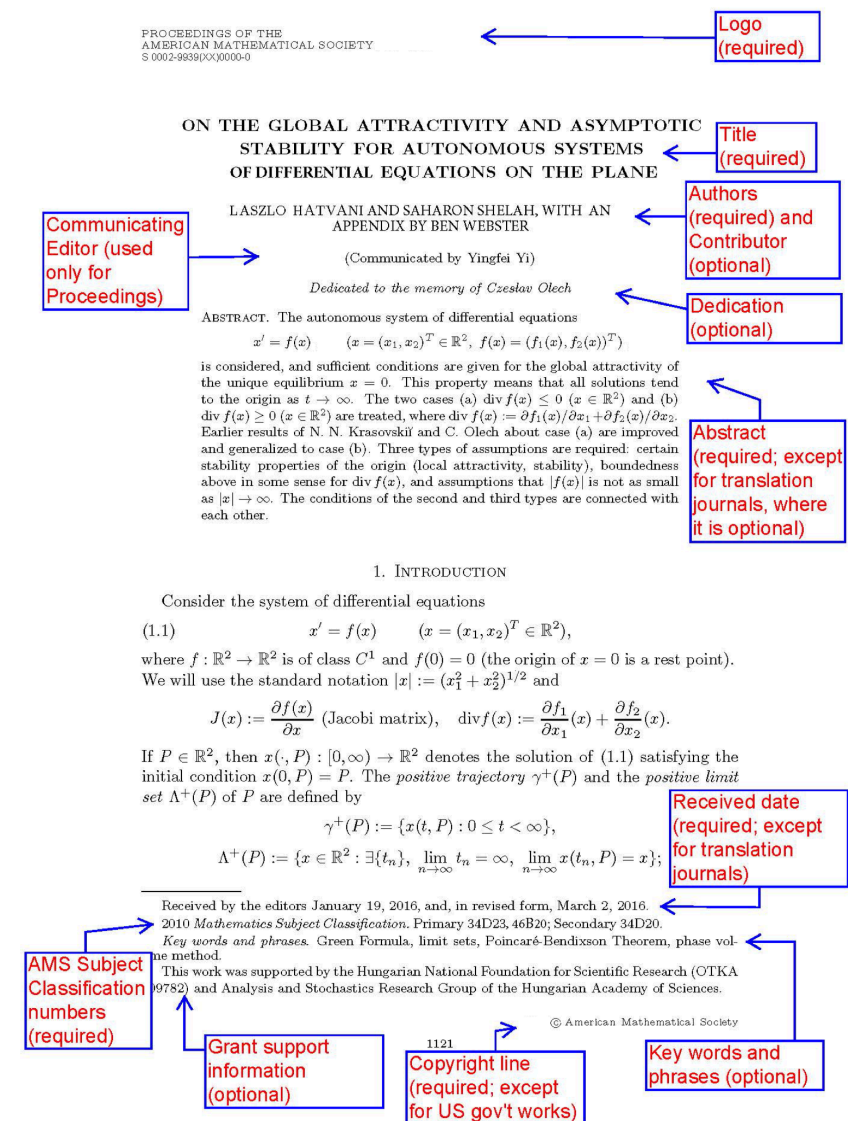
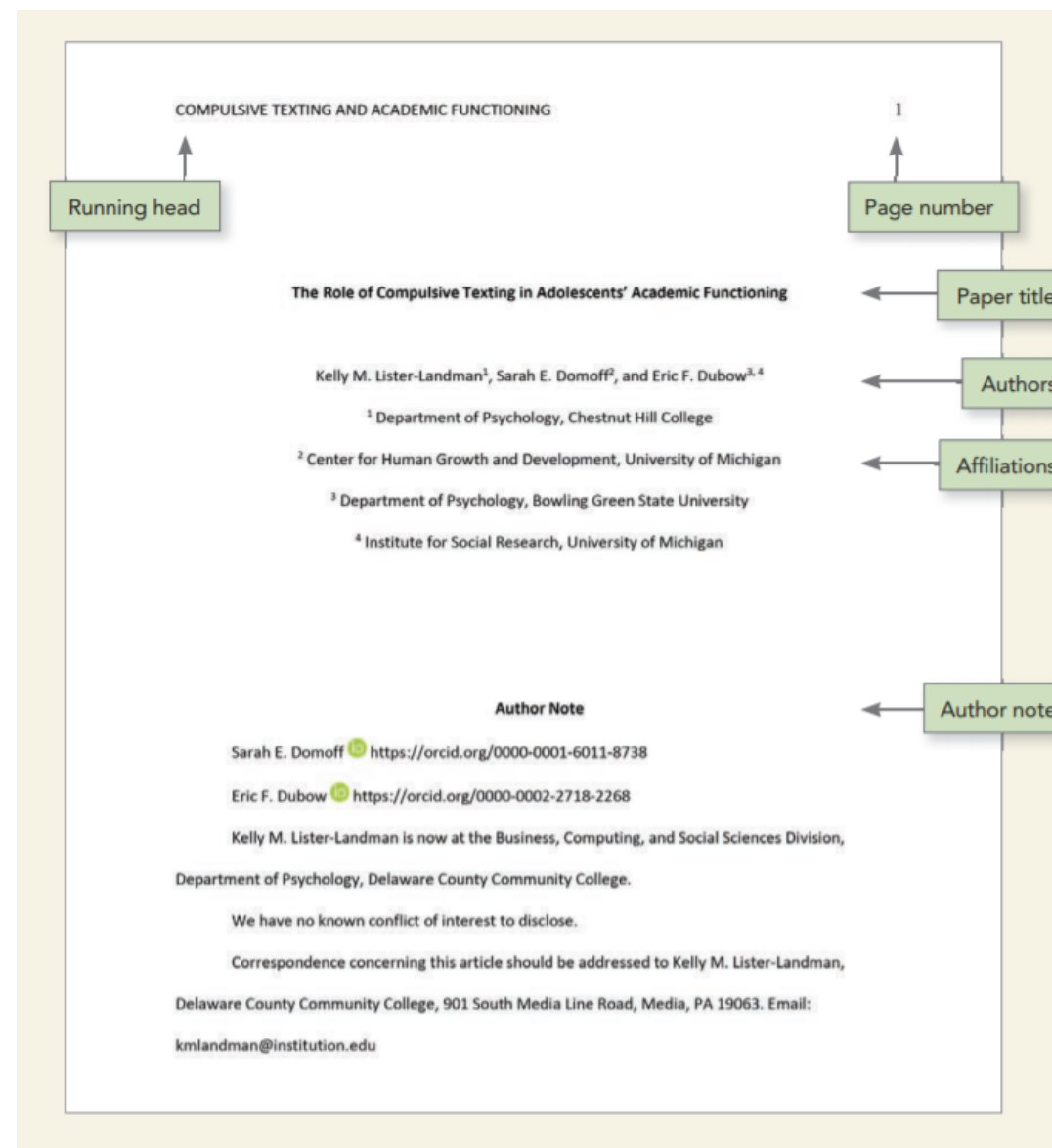
- >> Overview: *Reporting Statistics, Writing* (30 min)
- >> Hands-on Activity (20 min)
- >> Stats session (20 min)

## *What are reporting norms in HCI research?*

Because HCI is a rather eclectic field, the reporting norms are adopted from different fields, roughly as follows:

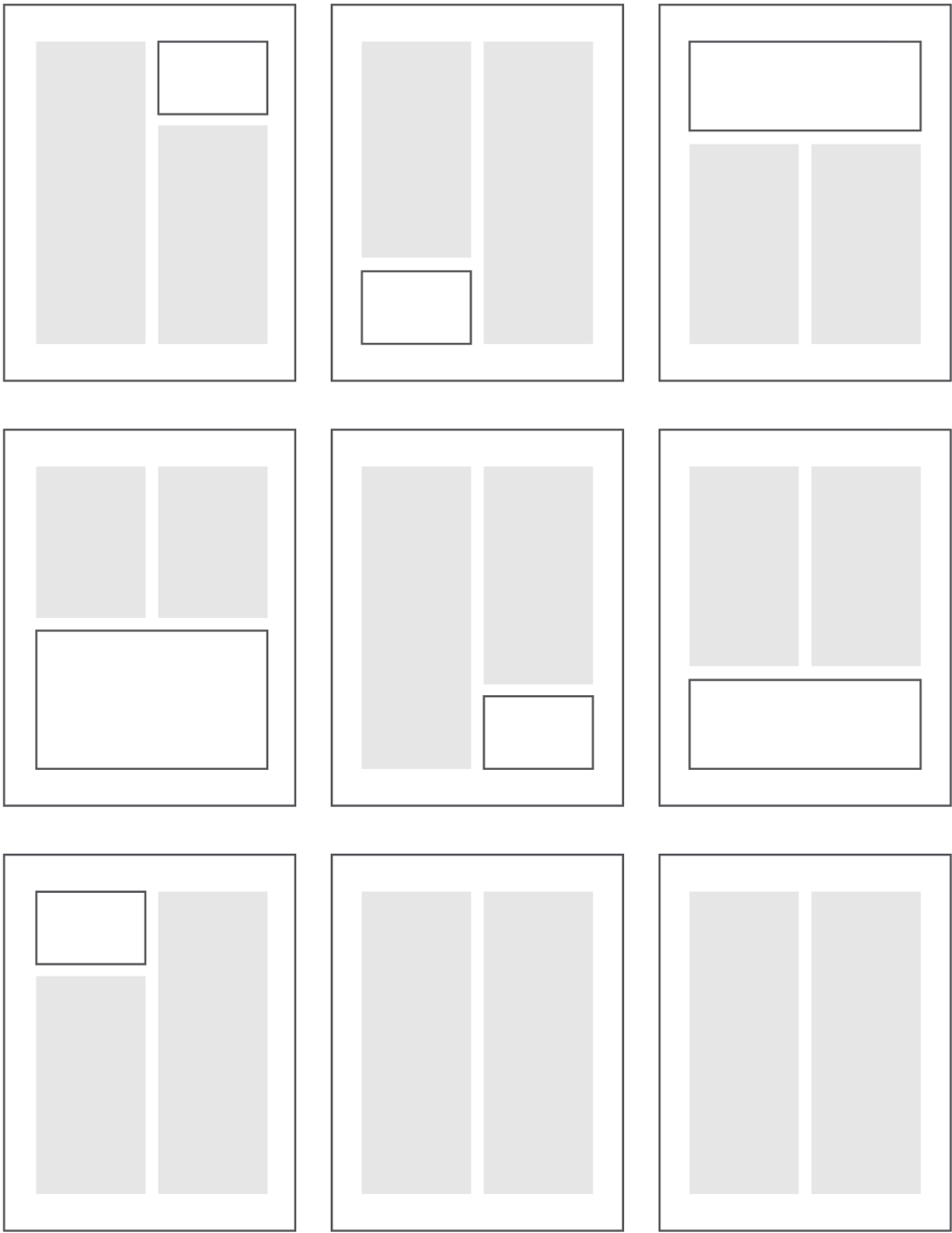
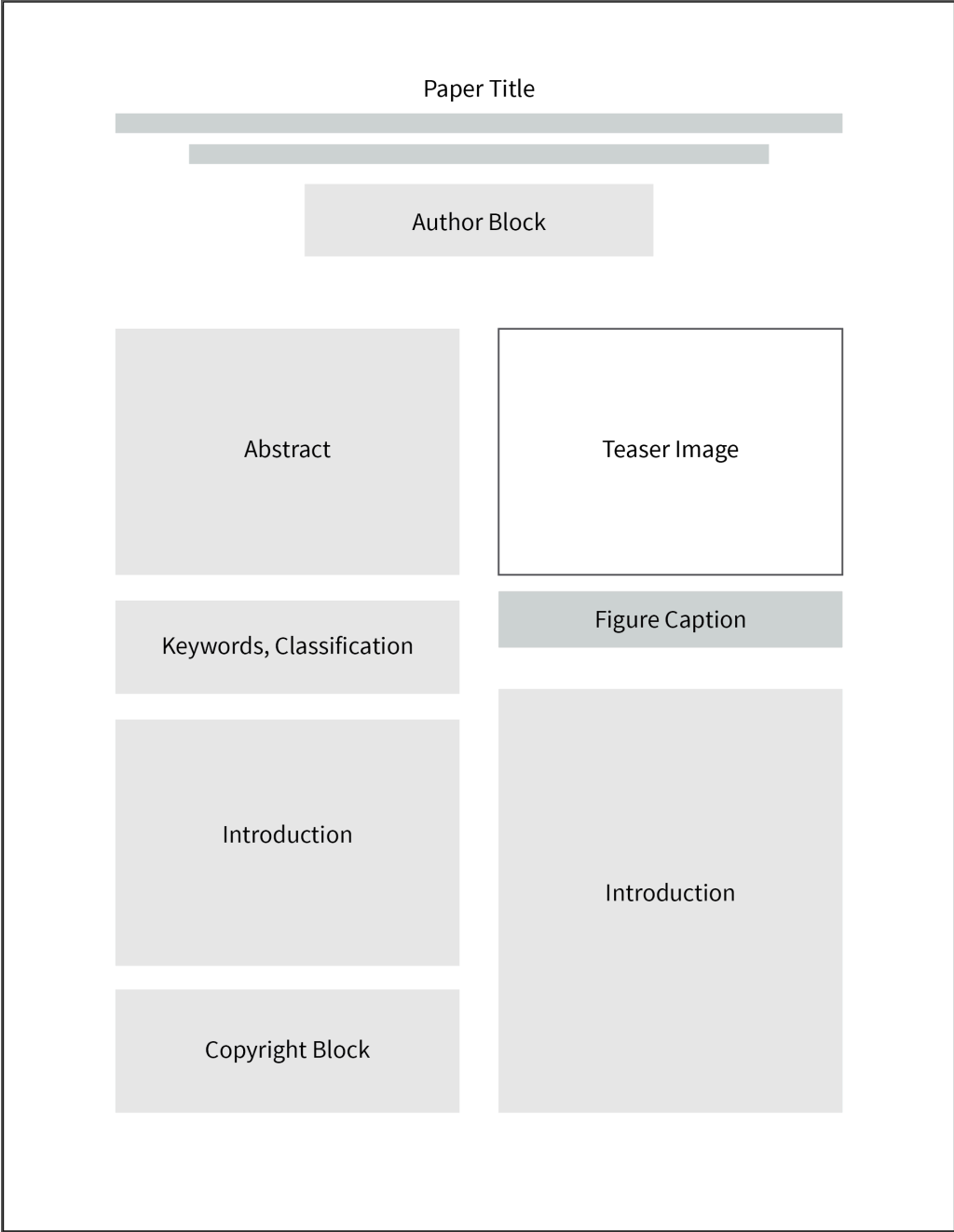
<b>Aspect</b>	<b>Norm</b>
Paper structure	APA (loosely)
Results of statistical analyses	APA (strictly)
Tables, figures	APA (very loosely)
Citations	Depends on the publisher (ACM, IEEE, etc.)
Formulas	AMS (loosely)
Style	APA (loosely), generally high standards in writing

# APA Publication Manual: Print, Web; AMS Style Guide: Web<sup>1</sup>



<sup>1</sup> Sources: Left, Right

# What does an HCI paper look like?



## *How is an HCI paper structured?*

HCI papers commonly follow the structure below:

- >> Abstract
- >> Introduction
- >> Related Work/Background
- >> *Hypotheses (quant. empirical)*
- >> *System/Design (design-based)*
- >> Method
- >> Results
- >> Discussion
- >> Conclusion
- >> Acknowledgements
- >> References
- >> Appendices

## *What is an abstract?*<sup>2</sup>

The abstract provides a brief but comprehensive summary of the contents of the paper. It gives readers an overview of the paper and helps them decide whether to read the full text. Usually *150 words* max.

The abstract usually includes (1–2 sentences each):

- >> Summary of literature review
- >> Problem investigated/RQs
- >> hypotheses
- >> methods used
- >> study results
- >> implications

---

<sup>2</sup> APA



## *How do I choose a title?*

There is no formula or requirement, but a few things to consider:

- >> It should be as short as it can be, but not too broad.
  - >> E.g., *Bodystorming Human-Robot Interactions*
- >> A common format in HCI:
  - >> Catchy headline/System name: Technical title
  - >> E.g., *Pay attention!: Designing adaptive agents that monitor and improve user engagement*
  - >> E.g., *Reading socially: Transforming the in-home reading experience with a learning-companion robot*

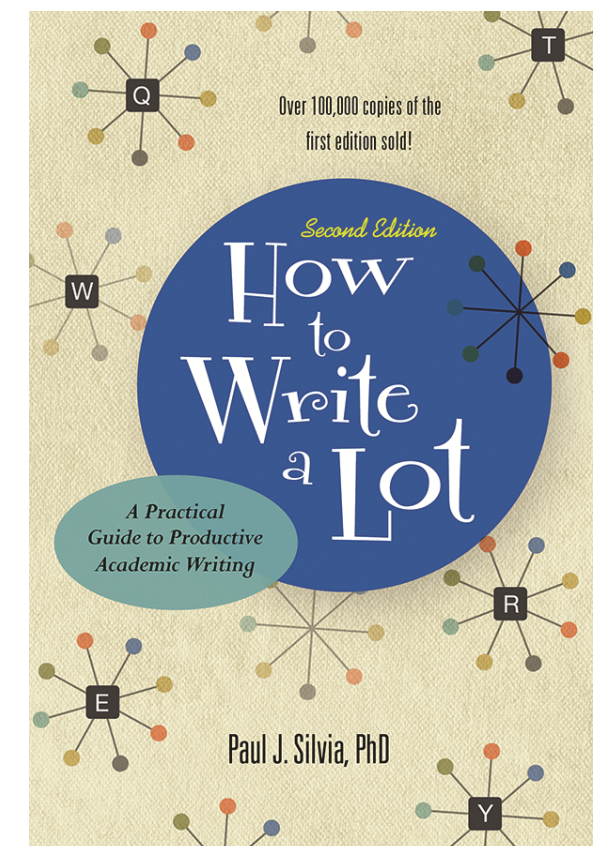
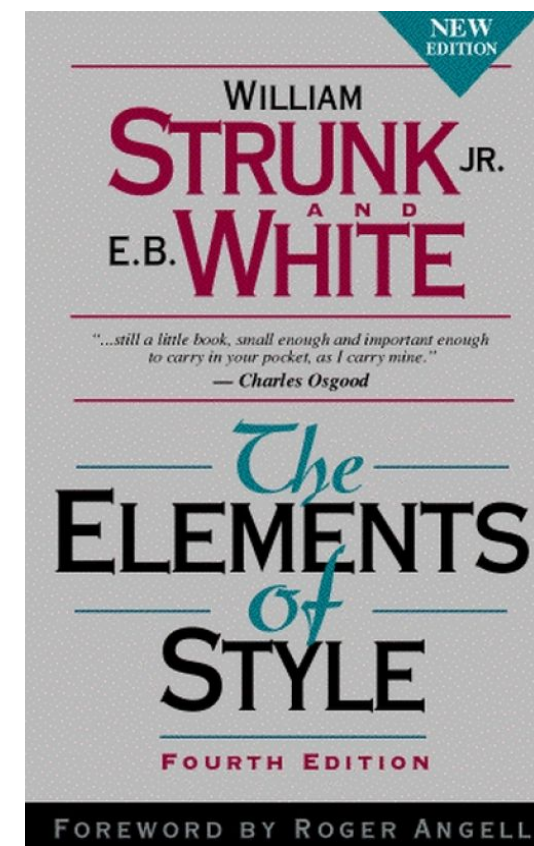
*What are other things I should pay attention to?*

1. Writing
2. Formatting
3. Presentation

## Writing<sup>3</sup>

The HCI community pays more attention to writing than most other CS communities, so writing is very important, in particular:

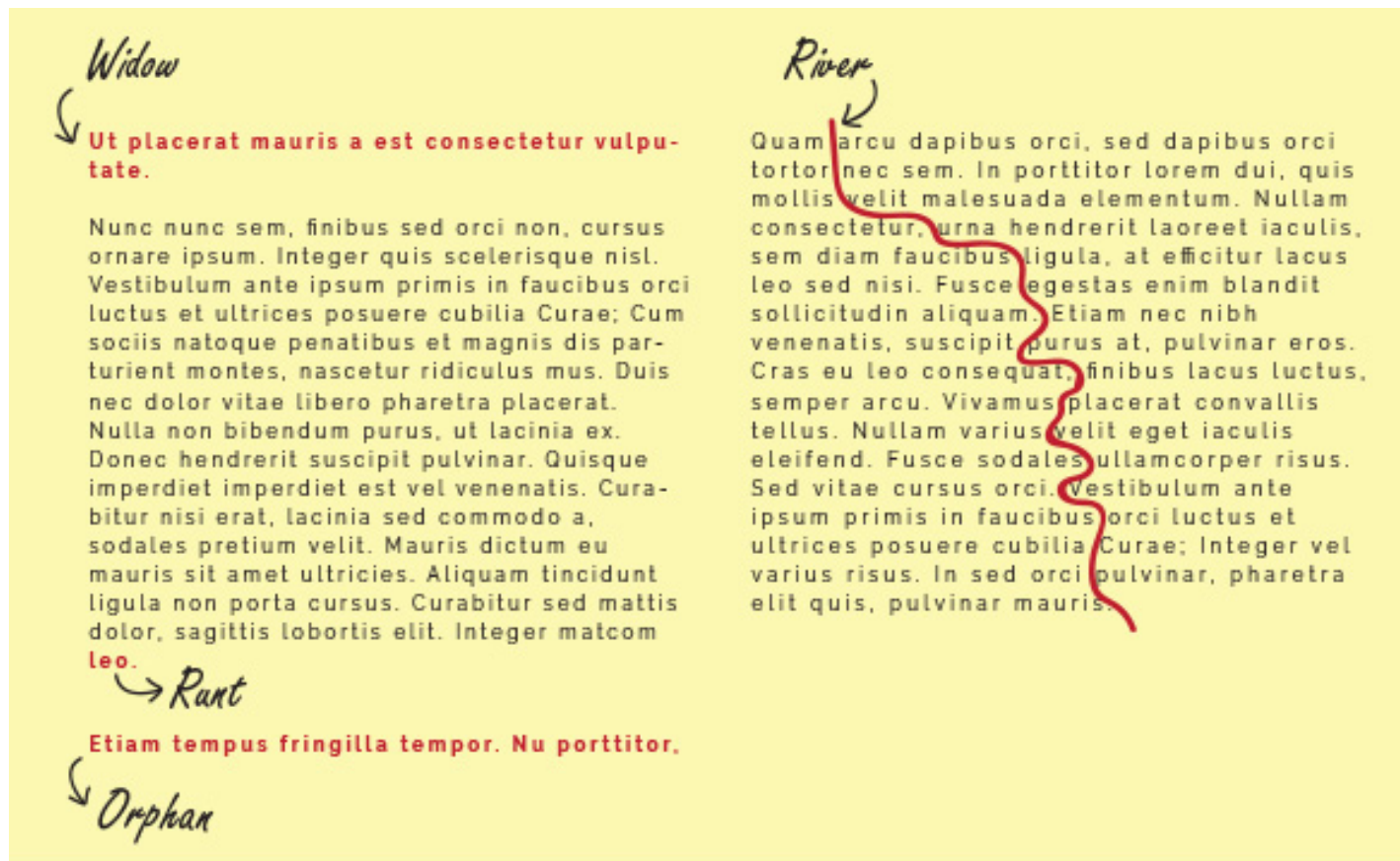
1. Reporting as *storytelling*
2. Flow among parts
3. "Cut deadwood"
4. Avoid any deviation from rules (syntax, grammar, punctuation, etc.)



<sup>3</sup>Image sources: [Left](#), [Right](#)

## Formatting<sup>4</sup>

For good *typography*, become familiar with *leading*, *tracking*, *kerning*, *widows*, *orphans*, *runts*, *rags*, *rivers*.



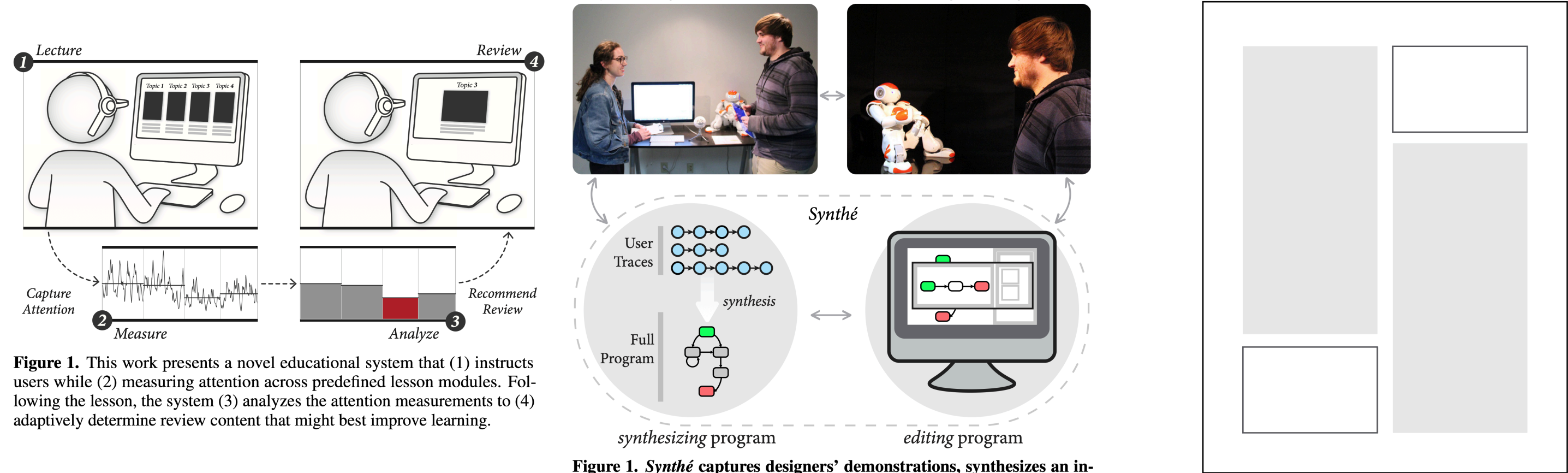
<sup>4</sup>Image source: [Left](#), [Right](#)

kerning  
tracking  
leading  
point size  
typeface  
justification  
line width

what is typography?

# Presentation<sup>5</sup>

The overall organization and visual appearance, using informative figures (e.g., a "teaser"), will improve accessibility and appeal.



**Figure 1.** This work presents a novel educational system that (1) instructs users while (2) measuring attention across predefined lesson modules. Following the lesson, the system (3) analyzes the attention measurements to (4) adaptively determine review content that might best improve learning.

**Figure 1.** Synthé captures designers' demonstrations, synthesizes an in-

<sup>5</sup> **Left:** Szafir & Mutlu, 2014; **Center:** Porfirio et al., 2019

*How do we report statistics?*

**Descriptive statistics:** Distribution characteristics using summary statistics in text, tables, or graphs.

**Inferential statistics:** Test parameters and results in text or tables and highlighting of significance in graphs.

In *text*, APA guidelines are strictly followed; in *graphs*, you can be creative.

# Descriptive statistics<sup>6</sup>

```
> describeBy(data$Guesses, list(data$Leakage,data$TBI))

Descriptive statistics by group
: Leakage
: HC
  vars  n mean  sd median trimmed  mad min max range skew kurtosis  se
X1    1 291 3.87 1.91     4    3.68 1.48  1 13   12 1.08    1.95 0.11
-----
: No Leakage
: HC
  vars  n mean  sd median trimmed  mad min max range skew kurtosis  se
X1    1 367 4.02 1.85     4    3.86 1.48  1 11   10 0.82    0.83 0.1
-----
: Leakage
: TBI
  vars  n mean  sd median trimmed  mad min max range skew kurtosis  se
X1    1 282 3.92 2.24     4    3.63 1.48  1 17   16 2.11    7.83 0.13
-----
: No Leakage
: TBI
  vars  n mean  sd median trimmed  mad min max range skew kurtosis  se
X1    1 353 4.37 2.46     4    4.05 1.48  1 19   18 1.55    4.24 0.13
```

The healthy controls guessed the item that the robot picked in 3.97 guesses ( $SD=1.91$ ) when the robot gazed toward the item and in 4.02 guesses ( $SD=1.85$ ) when the robot did not gaze toward it. Participants with TBI guessed the robot's pick in 3.92 guesses ( $SD=2.24$ ) when the robot gazed toward it and in 4.37 guesses ( $SD=2.46$ ) when the robot did not.

---

<sup>6</sup>Data from [Mutlu et al., 2018, Social-cue perception](#)

## How do we deal with decimals?<sup>7</sup>

For numbers...	Round to...	SPSS	Report
Greater than 100	Whole number	1034.963	1035
10 - 100	1 decimal place	11.4378	11.4
0.10 - 10	2 decimal places	4.3682	4.37
0.001 - 0.10	3 decimal places	0.0352	0.035
Less than 0.001	As many digits as needed for non-zero	0.00038	0.0004

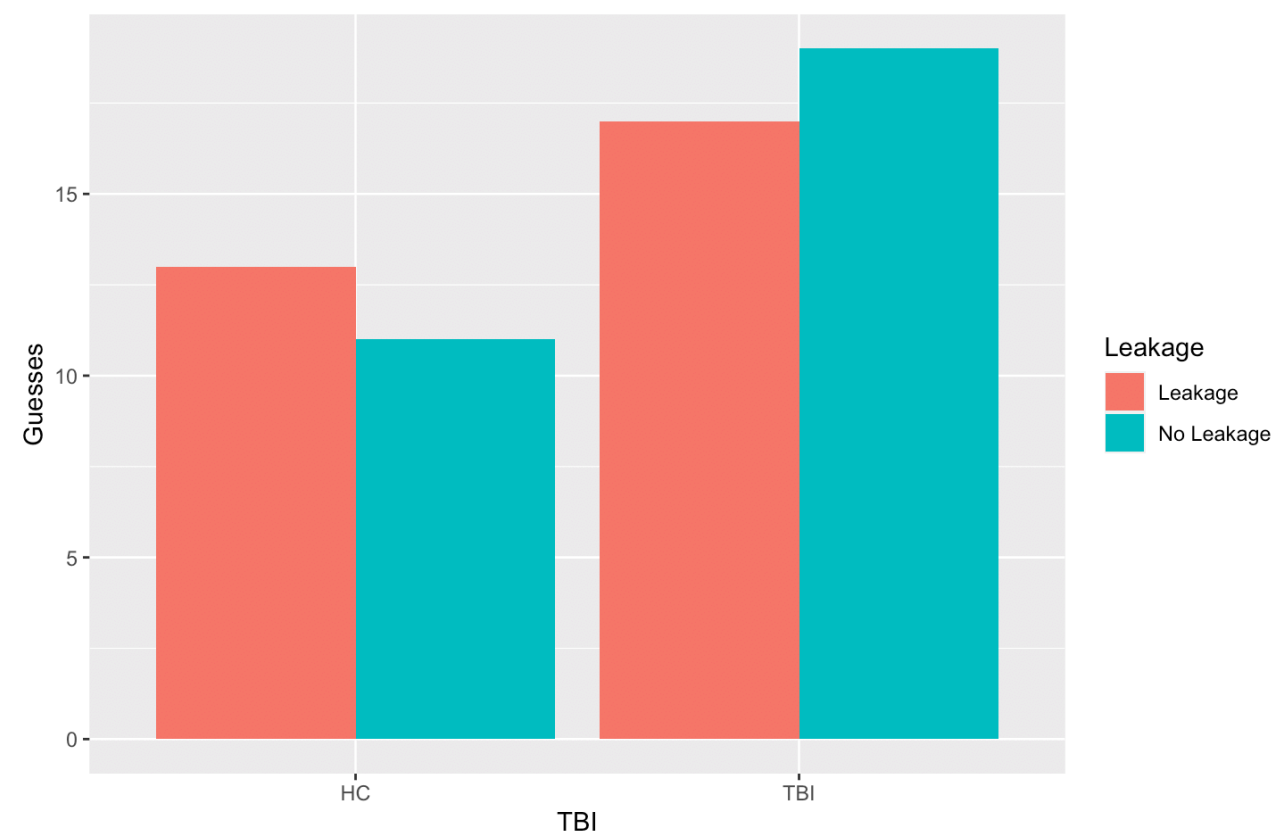
---

<sup>7</sup>Source



# Descriptive statistics (visual)<sup>8</sup>

```
library(ggplot2)  
ggplot(data, aes(fill=Leakage, y=Guesses, x=TBI)) +  
  geom_bar(position="dodge", stat="identity")
```



<sup>8</sup> [More information on using ggplot2](#)

# Inferential statistics<sup>9</sup>

```
> summary(aov(Guesses~(TBI*Leakage)+Error(ID/Leakage)+TBI,data=data))
```

Error: ID

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
TBI	1	15.2	15.236	2.360	0.127
Leakage	1	4.0	4.012	0.621	0.432
TBI:Leakage	1	7.5	7.467	1.157	0.284
Residuals	142	916.6	6.455		

Error: ID:Leakage

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Leakage	1	27.3	27.268	6.680	0.0107 *
TBI:Leakage	1	7.1	7.131	1.747	0.1884
Residuals	144	587.8	4.082		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Error: Within

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Residuals	1001	4325	4.321		

A mixed-model analysis of variance (ANOVA) revealed a significant effect of the leakage cue,  $F(1,144) = 6.68, p = .011$ .

Participants correctly identified the robot's pick on an average of 3.89 questions ( $SD = 2.08$ ) when the robot displayed the gaze cue and 4.19 ( $SD = 2.17$ ) when it did not.

---

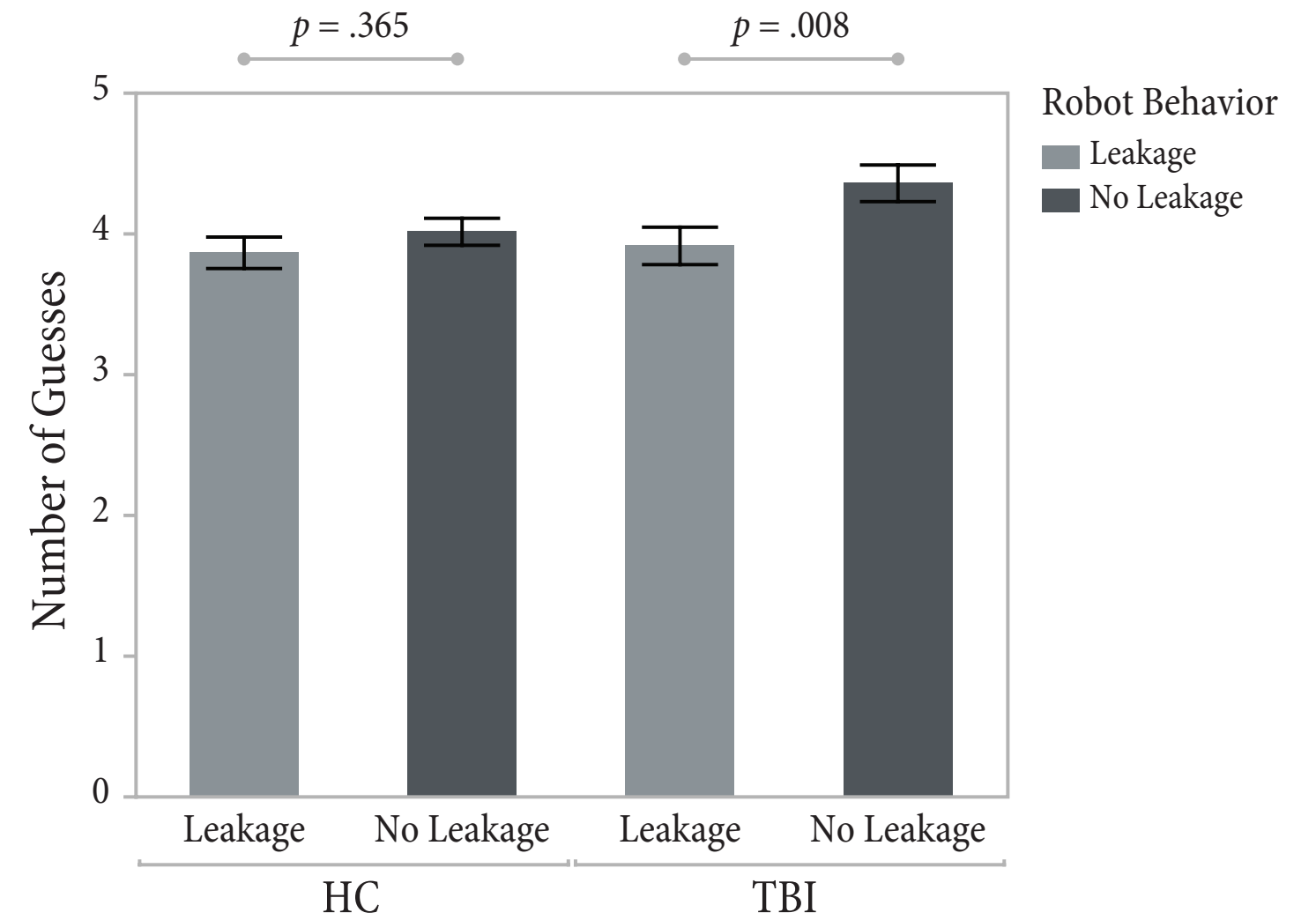
<sup>9</sup> Shown is a simplified model using data from [Mutlu et al., 2018](#)

## How do I report different tests?<sup>7</sup>

Statistic	Example
Mean and standard deviation	$M = 3.45, SD = 1.21$
Mann-Whitney	$U = 67.5, p = .034, r = .38$
Wilcoxon signed-ranks	$Z = 4.21, p < .001$
Sign test	$Z = 3.47, p = .001$
t-test	$t(19) = 2.45, p = .031, d = 0.54$
ANOVA	$F(2, 1279) = 6.15, p = .002, \eta_p^2 = 0.010$
Pearson's correlation	$r(1282) = .13, p < .001$

<sup>7</sup>Source

Test results can also be mapped on graphs either manually (e.g., using Adobe Illustrator) or automatically using advanced scripting (e.g., ggplot2, matplotlib).



# Data Visualization with ggplot2 : : CHEAT SHEET



## Basics

ggplot2 is based on the **grammar of graphics**, the idea that you can build every graph from the same components: a **data set**, a **coordinate system**, and **geoms**—visual marks that represent data points.

To display values, map variables in the data to visual properties of the geom (**aesthetics**) like **size**, **color**, and **x** and **y** locations.

Complete the template below to build a graph.

```
ggplot(data = DATA) +
  GEOM_FUNCTION(mapping = AEST) +
  COORDINATE_FUNCTION(position = POSITION) +
  SCALE_FUNCTION +
  THEME_FUNCTION
```

**ggplot**(data = mpg, aes(x = cyl, y = hwy)) Begins a plot that you finish by adding layers. Add one geom function per layer.

**aes**(x = cyl, y = hwy, data = mpg, geom = "point") Creates a complete plot with given data, geom, and mappings. Supplies many useful defaults.

**last\_plot()** Returns the last plot

**ggsave**("plot.png", width = 5, height = 5) Saves last plot as 5" x 5" file named "plot.png" in working directory. Matches file type to file extension.

## Geoms

Use a geom function to represent data points, use the geom's aesthetic properties to represent variables. Each function returns a layer.

**GRAPHICAL PRIMITIVES**

a = **geom\_blank()** (Useful for expanding limits)

b = **geom\_curve**(aes(xend = lat + 1, yend = long + 1, curvature = 1), aes(x = lat, y = long, color = "red", linetype = "dashed", size = 2))

a = **geom\_path**(aes(x = "butt", y = "round", linetype = "solid"))

a = **geom\_polygon**(aes(x = group, y = group))

b = **geom\_rect**(aes(xmin = long, ymin = lat, xmax = long + 1, ymax = lat + 1))

a = **geom\_ribbon**(aes(x = unemploy - 500, y = unemploy + 500))

**LINE SEGMENTS**

b = **geom\_abline**(aes(intercept = 0, slope = 1))

b = **geom\_hline**(aes(yintercept = 10))

b = **geom\_vline**(aes(xintercept = 10))

b = **geom\_segment**(aes(x1 = 1, x2 = 2, y = 1))

b = **geom\_spoke**(aes(x = 1, y = 1, radius = 1))

**ONE VARIABLE continuous**

c = **geom\_area**(aes(x = "bin", y = count))

c = **geom\_density**(aes(x = "bin", y = density))

c = **geom\_dotplot**(aes(x = "bin", y = count))

c = **geom\_freqpoly**(aes(x = "bin", y = count))

c = **geom\_histogram**(aes(x = "bin", y = count))

c2 = **geom\_qq**(aes(x = "bin", y = count))

**discrete**

d = **geom\_bar**(aes(x = "bin", y = count))

## TWO VARIABLES

**continuous x, continuous y**

e = **geom\_label**(aes(x = cyl, y = hwy))

e = **geom\_jitter**(aes(x = cyl, y = hwy))

e = **geom\_point**(aes(x = cyl, y = hwy))

e = **geom\_quantile**(aes(x = cyl, y = hwy))

e = **geom\_rug**(aes(x = cyl, y = hwy))

e = **geom\_smooth**(aes(x = cyl, y = hwy))

e = **geom\_text**(aes(x = cyl, y = hwy))

**discrete x, continuous y**

f = **geom\_col**(aes(x = "bin", y = count))

f = **geom\_boxplot**(aes(x = "bin", y = count))

f = **geom\_dotplot**(aes(x = "bin", y = count))

f = **geom\_violin**(aes(x = "bin", y = count))

**discrete x, discrete y**

g = **geom\_count**(aes(x = "bin", y = count))

## THREE VARIABLES

**continuous bivariate distribution**

h = **geom\_bin2d**(aes(x = "bin", y = count))

h = **geom\_density2d**(aes(x = "bin", y = count))

h = **geom\_hex**(aes(x = "bin", y = count))

**continuous function**

i = **geom\_area**(aes(x = "bin", y = count))

i = **geom\_line**(aes(x = "bin", y = count))

i = **geom\_step**(aes(x = "bin", y = count))

**visualizing error**

j = **geom\_crossbar**(aes(x = "bin", y = count))

j = **geom\_errorbar**(aes(x = "bin", y = count))

j = **geom\_linerange**(aes(x = "bin", y = count))

j = **geom\_pointrange**(aes(x = "bin", y = count))

**maps**

k = **geom\_map**(aes(x = "bin", y = count))

k = **geom\_raster**(aes(x = "bin", y = count))

k = **geom\_tile**(aes(x = "bin", y = count))

## Stats

An alternative way to build a layer

A stat builds new variables to plot (e.g., count, prop).

data | stat | geom | coordinate system

Visualize a stat by changing the default stat of a geom function, **geom\_stat**(stat="count") or by using a stat function, **stat\_count**(geom="bar"), which calls a default geom to make a layer (equivalent to a geom function). Use **..name..** syntax to map stat variables to aesthetics.

geom to use | stat function | geom mapping

i = **stat\_density2d**(aes(fill = ..level..), geom = "polygon")

c = **stat\_bin**(bins = 30, drop = T)

c = **stat\_bin\_hex**(bins = 30, drop = T)

c = **stat\_bin\_2d**(bins = 30, drop = T)

c = **stat\_contour**(aes(x = "bin", y = count))

c = **stat\_boxplot**(coef = 1.5)

c = **stat\_smooth**(method = "lm", formula = y ~ x, level = 0.95)

ggplot() = **stat\_function**(aes(x = -3, y = 99, fun = dnorm, args = list(0.5, 1)), aes(x = ..x.., y = ..y..))

ggplot() = **stat\_qq**(aes(sample = 1:100, dist = qt, qnorm = list(0.5, 1)), aes(x = ..sample.., y = ..qnorm..))

e = **stat\_summary**(fun.data = "mean", geom = "bar")

e = **stat\_unique()**

## Scales

Scales map data values to the visual values of an aesthetic. To change a mapping, add a new scale.

scale = aesthetic to adjust | prepackaged scale to use | scale specific arguments

range of values to include | ticks to use in legend axis | ticks to use in legend axis

**GENERAL PURPOSE SCALES**

scale\_\*\_continuous() - map cont values to visual ones

scale\_\*\_discrete() - map discrete values to visual ones

scale\_\*\_manual() - use data values as visual ones

scale\_\*\_manual(values = c("red", "blue", "yellow"))

scale\_\*\_manual(values = c("red", "blue", "yellow"), name = "sex", labels = c("M", "F", "U"))

**X & Y LOCATION SCALES**

scale\_x\_log10() - Plot x on log10 scale

scale\_x\_reverse() - Reverse direction of x axis

scale\_x\_sqrt() - Plot x on square root scale

**COLOR AND FILL SCALES (DISCRETE)**

a = **scale\_fill\_brewer**(palette = "Blues")

a = **scale\_fill\_grey**(start = 0.2, end = 0.8, na.value = "red")

**COLOR AND FILL SCALES (CONTINUOUS)**

a = **scale\_color\_distill**(palette = "Blues")

a = **scale\_fill\_gradient**(low = "red", high = "yellow")

a = **scale\_fill\_gradient2**(low = "red", high = "blue", mid = "white", midpoint = 25)

a = **scale\_fill\_gradientn**(colors = top.colors(10))

**SHAPE AND SIZE SCALES**

p = **scale\_shape\_manual**(values = c(1, 2))

p = **scale\_size\_area**(max\_size = 100)

## Coordinate Systems

r = **d + geom\_bar**

r = **coord\_cartesian**(xlim = c(0, 5), ylim = c(0, 10))

r = **coord\_fixed**(ratio = 1/1)

r = **coord\_flip**

r = **coord\_map**

r = **coord\_polar**(theta = "x", direction = 1)

r = **coord\_trans**(trans = "sqrt")

r = **coord\_quickmap**

**Position Adjustments**

s = **geom\_bar**(position = "dodge")

s = **geom\_point**(position = "jitter")

s = **geom\_label**(position = "nudge")

s = **geom\_bar**(position = "stack")

Each position adjustment can be recast as a function with manual width and height arguments

s = **geom\_bar**(position = position\_dodge(width = 1))

## Faceting

Facets divide a plot into subplots based on the values of one or more discrete variables.

t = **ggplot**(mpg, aes(x = cyl, y = hwy)) + **geom\_point**()

t = **facet\_grid**(~ hwy) | facet into columns based on hwy

t = **facet\_grid**(year ~ .) | facet into both rows and columns

t = **facet\_wrap**(~ hwy) | wrap facets into a rectangular layout

Set scales to let axis limits vary across facets

t = **facet\_grid**(drv ~ hwy, scales = "free") | x and y axis limits adjust to individual facets

t = **facet\_grid**(~ hwy, label = label\_both) | label both axes

t = **facet\_grid**(~ hwy, label = label\_both, labs(x = "x", y = "y")) | label both axes

t = **facet\_grid**(~ hwy, label = label\_parsed) | parse labels

Labels

l = **labs**(x = "New x axis label", y = "New y axis label", title = "Add a title above the plot", subtitle = "Add a subtitle below title", caption = "Add a caption below plot", notes = "New notes legend title")

l = **annotate**(geom = "text", x = 8, y = 9, label = "X")

Legends

a = **theme**(legend.position = "bottom")

a = **guides**(fill = "none")

a = **scale\_fill\_discrete**(name = "Title", labels = c("A", "B", "C"), breaks = c("A", "B", "C"))

Zooming

z = **coord\_cartesian**(xlim = c(10, 20), ylim = c(10, 20))

z = **xlim**(10, 20) + **ylim**(10, 20)

z = **scale\_x\_continuous**(limits = c(0, 100)) + **scale\_y\_continuous**(limits = c(0, 100))

